AD-A042 685 TEXAS UNIV AT AUSTIN DEPT OF ELECTRICAL ENGINEERING F/G 12/1
THE STUDY OF DISTRIBUTION-FREE PERFORMANCE BOUNDS FOR NONPARAME--ETC(U)
JUN 77 T J WAGNER AF-AFOSR-2371-72

UNCLASSIFIED AFOSR-TR-77-0879 NL

1 OF 1

ADA042 685

END

DATE FILMED

8 - 77

DDC

FINAL REPORT FOR

AF-AFOSR ~~Grant~~ 72-2371-72

THE STUDY OF DISTRIBUTION-FREE PERFORMANCE BOUNDS FOR NONPARAMETRIC DISCRIMINATION ALGORITHMS; Final Report AFOSR Grant 72-2371; 6/20/77 by T.J. Wagner

T. J. Wagner

20 Jun 77

INTRODUCTION.

14 p.

The accomplishments of the work performed with whole or partial support of the Grant AFOSR 72-2371 are outlined in Section I. The papers published with grant support are given in Section II along with those that are in some prepublication stage. Section III lists the technical personnel supported by the grant during each year period.

401 997

SECTION I. ACCOMPLISHMENTS.

The main grant research effort has been in the area of nonparametric discrimination and the closely related problem of density estimation. In the discrimination problem a statistician makes an <u>observation</u> X, a random vector with values in $\mathbb{R}^d$, and wishes to estimate its <u>state</u> θ, a random variable taking values in $\{1,\ldots,M\}$. All that he knows about the distribution of (X,θ) is that which can be inferred from a sample (X_1,θ_1), $\ldots,(X_n,\theta_n)$ of size n drawn from that distribution. The sample, commonly called <u>data</u>, is assumed to be independent of (X,θ). Using X and the data the statistician makes a randomized decision $\hat{\theta}$ for θ where his <u>rule</u> is any procedure which determines the probability distribution $\delta = (\delta_1,\ldots,\delta_M)$ for $\hat{\theta}$ given X and the data. In particular,

$$\delta: \ \mathbb{R}^d \times (\mathbb{R}^d \times \{1,\ldots,M\})^n \to [0,1]^M \ ,$$

$$\sum_1^M \delta_j = 1 \ , \text{ and}$$

$$P\{\hat{\theta} = j \mid X,(X_1,\theta_1),\ldots,(X_n,\theta_n)\} = \delta_j \ , \ 1 \le j \le M \ .$$

For his data and rule, the probability of error is

$$L_n = P\{\hat{\theta} \neq \theta \mid (X_1,\theta_1),\ldots,(X_n,\theta_n)\} \ ,$$

a random variable whose value is the limiting frequency of errors made when a large number of independent observations have their states estimates with δ and the given data.

The questions which have been considered in the past usually are of the asymptotic type. For example, what does L_n converge to as n tends to infinity and if, say, $L_n \overset{n}{\to} L$ in probability, how does L compare to the Bayes probability of error L^*. Other asymptotic studies have been concerned with how one estimates L, or L^* if it is different from L, from the data. Investigators have also been recently concerned with estimating L_n from the data, sometimes called error estimation. If $\hat{L}_n$ is a particular estimate of L_n the statistician would like to know how small

$$P[|\hat{L}_n - L_n| \geq \varepsilon] \tag{1}$$

is for a given $\varepsilon > 0$. The difficulty is that the distribution of (X,θ) is unknown and the best that one can hope for is a distribution-free upper-bound for (1) which depends only on n, ε and d and which tends to zero with n for each fixed ε and d.

With grant support, a summary of all recent work on nonparametric discrimination, including density estimation, was completed [12]. In [20, 41] a class of rules, which are a natural generalization of k-nearest neighbor rules and which are called voting rules, are discussed and conditions are given for which $L_n \overset{n}{\to} L^*$ in probability and with probability one. In [19, 43], the asymptotic properties of k-nearest neighbor rules are investigated for the nonparametric estimation problem (θ now takes values in $\mathbb{R}^p$ rather than in $\{1,\ldots,M\}$).

A consideration always present in nonparametric discrimination is how to implement rules derived from large amounts of data. For example, if one uses the k-nearest neighbor rule with the data, a large n presents difficulties in that both storage requirements and computation times increase with n. In order to keep the implementation requirements within reason and still retain the appeal of k-nearest neighbor rules, various procedures for condensing or editing the data before the k-nearest neighbor rule is applied have been suggested. In [5], a simple argument for the asymptotic property of one of the most appealing of these schemes is given while in [23], a basic flaw in the original argument used for obtaining the asymptotic properties of the edited nearest neighbor rule is pointed out.

In error estimation, the first distribution-free bound for (1) was given for k-local rules in [29]† where $\hat{L}_n$ is the deleted estimate of L_n. A rule is k-local if the decision $\hat{\theta}$ depends only on X and the pairs (X_i, θ_i) for which X_i is one of the k-closest to X from $X_1, \ldots, X_n$. The bound for (1) given in [29] is of the form $A/n\varepsilon^2$ where A is an explicitly given small constant depending only on M and k. *Thus deleted estimates appear* to be very good estimates of L_n when k-local rules are used. This has also been confirmed in the extensive simulation [39] which shows that the bound of [29] is pessimistic. The deleted estimate has also been shown to be an asymptotically consistent estimate of the Bayes risk L* for a wide variety of rules [2]. Using the resubstitution estimate of L_n, an exponential bound for (1) has been found for linear rules and condensed nearest neighbor rules [17,22,26,38]. Bounds of the type $A/\sqrt{n}$ have also been obtained for (1) for the popular class of two-step rules [42].

† Ironically, this first and, in some ways, best paper that I've been involved in for this area has had to undergo several revisions with long editorial delays between them. The third revision is now being looked at.

In nonparametric density estimation one seeks to estimate the density f of the sample $X_1,\ldots,X_n$. Two popular estimates have been investigated under this grant. The kernel estimate f_n is given by

$$f_n(x) = \sum_{i=1}^{n} K((x - X_i)/h_n)/nh_n^{\,d} \tag{2}$$

where K, the kernel, is a probability density on $\mathbb{R}^d$ and $\{h_n\}$ is a sequence of positive numbers tending to zero with n. In [33] conditions are given which insure that

$$f_n(x) \to f(x) \quad \text{w.p.1} \tag{3}$$

$$\sup_x |f_n(x) - f(x)| \to 0 \quad \text{w.p.1}\ . \tag{4}$$

In both cases, these conditions are weaker than any in the literature. Additionally, in [34] conditions are given which insure

$$\int_{\mathbb{R}^d} |f_n(x) - f(x)|dx \to 0 \quad \text{w.p.1}\ . \tag{5}$$

Similar types of results are developed in [36]. The sequence $\{h_n\}$ in (2) is chosen without regard to $X_1,\ldots,X_n$, something which one would like to be able to do. In [11] conditions are given which allow $h_n = h_n(X_1,\ldots,X_n)$ while still yielding (3) and (4). These results are extended in [40].

The nearest neighbor estimate of f is given by

$$g_n(x) = \frac{k_n/n}{V_n}$$

where $\{k_n\}$ is a sequence of positive integers with $k_n \le n$ and

$$k_n/n \to \infty,$$

$$k_n \overset{n}{\to} \infty,$$

and V_n is the volume of the smallest sphere, centered at x, which contains k_n of the points $X_1,\ldots,X_n$. In [3] conditions are given for

$$g_n(x) \overset{n}{\to} f(x) \quad \text{w.p.1}$$

while in [22] conditions are given for

$$\sup_x |g_n(x) - f(x)| \to 0 \quad \text{w.p.1} .$$

Other results related to density estimation may be found in [4,31,32,35] while [27] contains a density estimation result which applies directly to the clustering problem.

Random search deals with the problem of locating the minimum of an unknown function g defined on $\mathbb{R}^d$. The function g is not assumed to possess any of the usual analytical properties, such as convexity, that are invoked when one is searching for extrema. For this purpose one generates a sequence $Z_1,Z_2,\ldots$ of random vectors with values in $\mathbb{R}^d$ and with Z_n representing the estimate of the location of a global minimum after n steps. The goal is to produce a sequence for which $g(Z_n)$ converges with probability one, to the essential infimum of $g(x)$. All random search methods generate a trial vector Z^*_{n+1} from $Z_1,\ldots,Z_n$ and then let

$$Z_{n+1} = \begin{cases} Z_n & \text{if } g(Z_n) < g(Z^*_{n+1}) \\ Z^*_{n+1} & \text{if } g(Z^*_{n+1}) \leq g(Z_n) . \end{cases}$$

A further complication can arise when, for any x, one cannot observe $g(x)$, but only a sample with distribution function G_x and mean $g(x)$.

If F_n denotes the distribution function of X^*_{n+1} given $X_1,\ldots,X_n$ and $\hat{g}(x)$ denotes the sample mean of a sample of a size λ_n with the distribution function G_x then Devroye [13], for

$$X_{n+1} = \begin{cases} X^*_{n+1} & \text{if } \hat{g}(X^*_{n+1}) < \hat{g}(X_n) - \varepsilon_n \\ X_n & \text{otherwise} , \end{cases}$$

gave conditions on F_n, $\{\varepsilon_n\}$, $\{\lambda_n\}$ which insure that

$$g(X_n) \overset{n}{\rightarrow} \text{ess inf } g(x) \quad \text{w.p.1} .$$

Further refinements may be found in [14,18,21,30,37].

The grant has also been used to partially support several applied pattern recognition projects. Under the direction of J.K. Aggarwal, a flying spot scanner system with color capabilities has been fabricated and used to analyze aerial color infrared photographs [24,28], primarily for the detection of diseases in citrus trees. Additionally, a computer analysis of planar curvilinear moving images was undertaken in [25].

Finally, the grant has supplied partial support for system theory work in [6,7], for the analysis and design of digital filters in [1,8,10, 15,16], and the analysis of computer storage systems in [9].

SECTION II. PAPERS PUBLISHED WITH AFOSR GRANT 72-2371 SUPPORT.

If a paper was published in the proceedings of a conference and later in a refereed journal, only the listing for the journal publication is given below.

[1] E.P.F. Kan and J.K. Aggarwal, "Multirate digital filtering," *IEEE Transactions on Audio and Electroacoustics*, AU-20, 223-225, 1972.

[2] T.J. Wagner, "Deleted estimates of the Bayes risk," *Annals of Statistics*, 1, 359-362, 1973.

[3] T.J. Wagner, "Strong consistency of a nonparametric estimate of a density function," *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3, 289-290, 1973.

[4] T.J. Wagner and C.K. Chow, "Consistency of an estimate of tree-dependent probability distributions," *IEEE Transactions on Information Theory*, IT-19, 369-371, 1973.

[5] T.J. Wagner, "Convergence of the edited nearest neighbor," *IEEE Transactions on Information Theory*, IT-19, 696-697, 1973.

[6] J.K. Aggarwal and D.H. Eller, "Optimization of functional differential systems," *Journal of Optimization Theory and Applications*, 11, 100-120, 1973.

[7] J.K. Aggarwal, "Feedback control of linear systems with distributed delay," *Automatica*, 9, 367-379, 1973.

[8] J.K. Aggarwal, "Input quantization and arithmetic round-off in digital filters - a review," *Network and Signal Theory* (J.K. Skwizynski and J.O. Scanlan, Eds.) Peter Peregrinus Ltd., London, 315-343, 1973.

[9] T.J. Wagner and P.A. Franaszek, "Some distribution-free aspects of paging algorithm performance," *Journal of the Association for Computing Machinery*, 21, 31-39, 1974.

[10] M.D. Ni and J.K. Aggarwal, "Two-dimensional filtering and its error analysis," *IEEE Transactions on Computers*, C-23, 942-954, 1974.

[11] T.J. Wagner, "Nonparametric estimates of probability densities," *IEEE Transactions on Information Theory*, IT-21, 438-440, 1975.

[12] T.M. Cover and T.J. Wagner, "Topics in statistical pattern recognition," *Digital Pattern Recognition*, K.S. Fu, Editor, Springer-Verlag, New York, 15-46, 1976.

[13] L.P. Devroye, "On the convergence of statistical search," *IEEE Transactions on Systems, Man and Cybernetics*, SMC-6, 46-56, 1976.

[14] L.P. Devroye, "Probabilistic search as a strategy selection procedure," *IEEE Transactions on Systems, Man and Cybernetics*, SMC-6, 315-321, 1976.

[15] M.D. Ni and J.K. Aggarwal, "Error analysis of two-dimensional recursive digital filters employing floating-point arithmetic," *IEEE Transactions on Computers*, C-25, 755-759, 1976.

[16] M.T. Manry and J.K. Aggarwal, "The design of multidimensional FIR digital filters by phase correction," *IEEE Transactions on Circuits and Systems*, CAS-23, 185-199, 1976.

[17] T.J. Wagner and L.P. Devroye, "A distribution-free performance bound in error estimation," *IEEE Transactions on Information Theory*, IT-22, 586-587, 1976.

[18] L.P. Devroye, "A class of optimal performance-directed probabilistic automata," *IEEE Transactions on Systems, Man and Cybernetics,* SMC-6, 777-784, 1976.

[19] C.S. Penrod, "Nonparametric estimation with local rules," Ph.D. Thesis, University of Texas at Austin, December 1976.

[20] L.P. Devroye, "Nonparametric discrimination and density estimation," Ph.D. Thesis, University of Texas at Austin, December 1976.

[21] L.P. Devroye, "On random search with a learning memory," *Proceedings of the International Conference on Cybernetics and Society,* Washington, D.C., November 1976.

[22] T.J. Wagner and L.P. Devroye, "The strong uniform consistency of the nearest neighbor density estimate," *Annals of Statistics,* 5, 536-540, 1977.

[23] T.J. Wagner and C.S. Penrod, "Another look at the edited nearest neighbor rule," *IEEE Transactions on Systems, Man and Cybernetics,* SMC-7, 92-94, 1977.

[24] S.A. Underwood and J.K. Aggarwal, "Interactive computer analysis of aerial color infrared photographs," *Computer Graphics and Image Processing,* 6, 1-24.

[25] W.K. Chow and J.K. Aggarwal, "Computer analysis of planar curvilinear moving images," *IEEE Transactions on Computers,* C-26, 179-185, 1977.

[26] L.P. Devroye and T.J. Wagner, "Distribution-free performance bounds with the resubstitution error estimate," *Proceedings of the IEEE Computer Society Conference on Pattern Recognition and Image Processing,* Troy, New York, 323-326, June 1977.

[27] L.P. Devroye and T.J. Wagner, "Asymptotic properties of clustering algorithms," *Proceedings of the IEEE Computer Society Conference on Pattern Recognition and Image Processing*, Troy, New York, 321-322, June 1977.

[28] M. Ali and J.K. Aggarwal, "Automatic detection and classification of infestations of crop insect pests and diseases from infrared aerial color photographs," *IEEE Transactions on Geoscience Electronics*, GE-15, 170-179, 1977.

Accepted Papers

[29] T.J. Wagner and W.H. Rogers, "A finite sample distribution-free performance bound for local discrimination rules," accepted by *Annals of Statistics*.

[30] L.P. Devroye, "An expanding automaton for use in stochastic optimization," to appear in *Cybernetics and Information Science*, K.S. Narendra, Editor.

[31] L.P. Devroye, "A uniform bound for the deviation of empirical distribution functions," accepted by the *Journal of Multivariate Analysis*.

[32] L.P. Devroye, "The uniform convergence of nearest neighbor regression function estimators and their application in optimization," accepted by the *IEEE Transactions on Information Theory*.

Papers Submitted

[33] L.P. Devroye and T.J. Wagner, "The strong uniform consistency of kernel density estimates," submitted to the *Annals of Statistics*.

[34] L.P. Devroye and T.J. Wagner, "The L_1 convergence of kernel density estimates, submitted to the *Annals of Statistics*.

[35] L.P. Devroye, "Simple recursive estimation of the mode of a multivariate density," submitted to the *Canadian Journal of Statistics.*

[36] L.P. Devroye, "On the pointwise and the integral convergence of recursive kernel estimates of probability densities," submitted to *Utilitas Mathematica.*

[37] L.P. Devroye, "The local convergence of probabilistic search algorithms," submitted to the *Journal of Optimization Theory and Applications.*

[38] T.J. Wagner and L.P. Devroye, "Distribution-free performance bounds with the resubstitution error estimate," submitted to the *IEEE Transactions on Information Theory.*

[39] T.J. Wagner and C.S. Penrod, "Risk estimation for nonparametric discrimination and estimation rules: a simulation study," submitted to the *IEEE Transactions on Systems, Man and Cybernetics.*

Papers in Preparation

[40] T.J. Wagner and L.P. Devroye, "Generalized kernel estimates."

[41] T.J. Wagner and L.P. Devroye, "Voting rules in nonparametric discrimination."

[42] T.J. Wagner and L.P. Devroye, "Error estimation for two-step rules."

[43] T.J. Wagner and C.S. Penrod, "Asymptotic properties of nearest neighbor rules in nonparametric estimation."

SECTION III. PERSONNEL SUPPORTED BY AFOSR 72-2371.

<u>6/1/72 - 5/31/73</u>

T.J. Wagner, Professor	3 summer months (full time)
J.K. Aggarwal, Professor	3 summer months (full time)
Baolian Liu, RA*	3 months (50% time)
D.J. Thompson, RA	8 months (22.5% time)

<u>6/1/73 - 5/31/74</u>

T.J. Wagner, Professor	3 summer months (full time)
J.K. Aggarwal, Professor	3 summer months (full time)

<u>6/1/74 - 5/31/75</u>

T.J. Wagner, Professor	3 summer months (full time)
J.K. Aggarwal, Professor	3 summer months (full time)
L.P. Devroye, RA	9 months (50% time)

<u>6/1/75 - 5/31/76</u>

T.J. Wagner, Professor	2 summer months (full time)
J.K. Aggarwal, Professor	1½ summer months (2/3 time)
L.P. Devroye, RA	7 months (50% time)
O. Teoh, RA	4½ months (25% time)
M. Day, RA	3½ months (12.5% time)

<u>6/1/76 - 5/31/77</u>

T.J. Wagner, Professor	2 summer months (full time)
L.P. Devroye, RA	7 months (50% time)
M. Ali, PRA**	3 months (full time)
M. Ali, PRA	2½ months (33.3% time)

* Research Assistant
** Post-doctoral Research Associate

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 77-0879	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE STUDY OF DISTRIBUTION-FREE PERFORMANCE BOUNDS FOR NONPARAMETRIC DISCRIMINATION ALGORITHMS		5. TYPE OF REPORT & PERIOD COVERED Final
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) T. J. Wagner		8. CONTRACT OR GRANT NUMBER(s) AFOSR 72-2371
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Texas, Department of Electrical Engineering, Austin, Texas 78712		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332		12. REPORT DATE 20 Jun 77
		13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

nonparametric discrimination, pattern recognition, clustering, density estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The accomplishments of the work performed with the whole or partial support of Grant AFOSR 72-2371, 6/1/72-5/31/77, are outlined along with a listing of the publications and technical personnel supported by the grant.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)